E7.3 10982
CR-133761

PROGRESS REPORT

June 20, 1973 - August 19, 1973

CROP IDENTIFICATION AND ACREAGE MEASUREMENT UTILIZING ERTS IMAGERY 013

PRINCIPLE INVESTIGATOR

DONALD H. VON STEEN AG 328

TABLE OF CONTENTS

i

(E73-10982) CROP IDENTIFICATION AND ACREAGE MEASUREMENT UTILIZING ERTS IMAGERY Progress Report, 20 Jun. - 19 Aug. 1973 (Department of Agriculture) 16 p HC $3.00 CSCL 02C G3/13

N73-31289

Unclas 00982

Segment Location

Progress

The locating of area segments and fields within segments has been the most time consuming phase of the Analysis. This laborious task has also been slowed by frequent shutdowns as well as slow turnaround time at the Washington Computer Center. The following table shows the status of this operation.

State	Number segments to locate	Number segments located	Number segments with fields located
Idaho.............	1/58	58	2/57
Kansas............	3/52	42	18
South Dakota......	50	12	0

1/ The total number of segments to locate was originally 62, however, 2 training segments had no useful ground truth data and 2 others were all fallow.
2/ In one segment the fields were too small to locate.
3/ There were originally 63 segments to locate, however, 5 have no useful ground truth data and 6 were covered with clouds.

Zoom Transfer Scope to Locate Segments

An attempt was made to locate segments from county maps onto ERTS imagery using the Zoom Transfer Scope at NASA. However, it was found that only small areas of two maps could be aligned and so only those segments with noticable features (a lake, river, etc.) nearby could be defined.

A possibility still exists that the Zoom scope would be useful in locating fields on the computer printout. Aircraft data would then be projected onto the ERTS printout and field boundaries traced. This procedure is one which will be tried since the tasks of delineating field boundaries is so slow and difficult.

Packaging

The new method of packaging ERTS false color transparencies in individual plastic packages is a welcome improvement. The package allows visual inspection of the imagery without removing the plastic covering and thus reduce wear and handling damage.

Microdensitometer

The microdensitometer was installed on August 1 by the Perkin-Elmer Corporation. During installation, it was discovered that the high speed reader/punch was missing. This missing item has not yet been received. Also, during installation it was noted that the apertures supplied were not the best suited for our use. A new aperture wheel will need to be ordered.

Acceptance tests are now being conducted. On August 10 the tape transport failed and to date the transport has only been partially repaired.

A 3-day training session followed the installation. The session covered such operating details as calibration, parameter entry, aperture settings and program loading. The functions of the software supplied were also demonstrated.

Data Processing

A program (program SRS 73048 - WMAP) has been developed with output comparable to the Penn State NMAP.

Advantages of using this program relative to NMAP are:

1. WMAP assigns class limits for a set of 9 printable characters based on a frequency distribution of a sample of a user defined area. These limits are set so that each character would appear about the

same number of times in a print-out of the sampled area. Therefore, individual maps from the sampled area should have reasonable contrast.

2. WMAP normally runs somewhat faster than NMAP. On the Washington Computer Center's 360-65, NMAP appears to map about 1,000 characters per second. WMAP averaged 1500 characters per second on a short (240 columns; 100 lines) run.
3. WMAP works from ERTS tapes, or from complete copies of ERTS tapes. Therefore, it is not limited by the dimensions of existing subsets.

Disadvantages are:

1. WMAP will map from response band 5 only.
2. The maximum width of any one map is 120 columns. If a wider scene is desired, an additional MAP AREA card will be required for each additional unit of up to 120 columns.

Missouri Crop Classification

Introduction

Personnel in the Statistical Reporting Service and LARS at Purdue University studied classification of 29 of the 50 area segments located in our Missouri test site. The analysis was performed on ERTS MSS system corrected tapes prepared by LARS. The data tapes used were geometrically corrected and temporally overlaid. The data for August 26 and October 2, 1972 were overlaid onto the September 14, 1972 data. Each ERTS tape for these dates contained 4 bands thus the overlaid LARS tape contained 12 bands. However, three of these 12 bands were of poor quality and could not be used.

Objectives

The objectives of this study are:

1. To estimate the Bayes Error 1/ rates of misclassification for three different discriminant analyses procedures, namely
   (a) Quadratic discriminant functions that assume equal prior probabilities.
   (b) Quadratic discriminant functions that use unequal prior probabilities, and
   (c) The per-field classification system in LARSYS.
2. To investigate the value of the temporal overlay, and
3. To estimate the classification matrix for Missouri on independent test data.

Procedures and Results

The first analysis was performed to compare three classification schemes. Two point discriminant analysis routines were run, one using equal prior probabilities, and one using probabilities proportional to the number of acres of each crop found in the segments in the test site.

Theory suggests that the discriminant functions that use unequal prior probabilities should be superior to the equal weighting model but previous experience has shown very little improvement. Since the ERTS tapes contained data for only 29 segments out of 50, we decided to use all data for both training and testing. The classification matrix will be biased because data are used for both purposes, however, this bias should be small if ample data are present.

1/ In this paper, Bayes Error means the optimum classification. It is a parametric procedure in which one assumes the distribution of the spectral data for each crop is known as well as the probability of each crop. Unknown spectral data are then classified as to the most likely crop.

The classification matrix (C.M.) obtained using quadratic discriminant functions which assume equal prior probabilities are presented in Table 1.

Table 1.--Classification matrix of quadratic discriminant functions with equal prior probabilities using data from 3 overflights 1/

Group	No. of sample points	Percent correct	Number of samples classified into					
			Cotton	Corn	Soybean	Grass	Winter wheat	Odd
Cotton........	927	74.3	689	21	83	36	61	37
Corn..........	58	58.6	4	34	3	10	5	2
Soybean.......	852	39.7	101	49	338	137	199	28
Grass.........	240	57.1	34	22	22	137	20	5
Winter wheat..	85	69.4	5	2	6	7	59	6
Odd...........	55	69.1	9	3	1	2	2	38
Totals........	2217		842	131	453	329	346	116

Overall performance 58.4

1/ August 26, 1972, MSS bands 4, 5, 7.
September 14, 1972, MSS bands 5, 7.
October 2, 1972, MSS bands 4, 5, 6, 7.

The next C.M. (Table 2) is for the quadratic discriminant functions with unequal prior probabilities. Prior probabilities can be computed several ways. One possibility is historic data, for example, last years' farm census. Another possibility is data from an earlier survey conducted during the same year. For this study, data collected in June was used as priors for the August 26 imagery. This is the most efficient but somewhat biased because the 1972 season's priors were obtained from the 1972 season's data. These weighted quadratic discriminant functions are the best possible point classifiers if the data are multivariate normal and the weights are nearly correct.

Table 2.--Classification matrix of quadratic discriminant functions with unequal prior probabilities using data from 3 overflights 1/

Group	No. of sample points	Percent correct	Number of samples classified into					
			Cotton	Corn	Soybean	Grass	Winter wheat	Odd
Cotton........	927	79.7	739	2	137	26	0	23
Corn..........	58	44.8	9	26	7	14	0	2
Soybean.......	852	71.8	99	12	612	96	8	25
Grass.........	240	53.3	42	1	66	128	0	3
Winter wheat..	85	22.4	9	1	40	10	19	6
Odd...........	55	70.9	8	1	4	3	0	39
Totals........	2217		906	43	866	277	27	98

Overall performance 70.5

1/ August 26, 1972, MSS bands 4, 5, 7.
September 14, 1972, MSS bands 5, 7.
October 2, 1972, MSS bands 4, 5, 6, 7.

A chi-square test for discriminatory power 2/ was run on the C.M. of Tables 1 and 2. The null hypothesis is that the classification was done strictly at random. If the null hypothesis is correct, then the spectral data were useless as far as giving information that would help assign the data to a crop class. If the above hypothesis is correct, then the statistic $\frac{(n-e)^2}{e} + \frac{(\bar{n}-\bar{e})^2}{\bar{e}}$ has a chi-square distribution with 1 degree of freedom. Where n and $\bar{n}$ are the number of correctly classified and misclassified points respectively and e and $\bar{e}$ are the expected number of correctly classified and misclassified points under the null hypothesis.

The chi-square for Table 1 is 2782 and for Table 2 is 4626. These chi-square values with one degree of freedom are highly significant and, therefore,

2/ S. James Press, "Applied Multivariate Analysis," Pages 381-383.

we conclude that the classification was not done at random. Another chi-square test based on the difference between the marginal sums and the correct number of data points in each class for Table 2 is as follows:

$$X^2_{(5)} = \frac{(906-927)^2}{927} + \frac{(43-58)^2}{58} + \frac{(866-852)^2}{852} + \frac{(277-240)^2}{240} + \frac{(27-85)^2}{85} +$$

$$\frac{(98-55)^2}{55} = .47 + 3.87 + .23 + 5.70 + 39.57 + 33.62 = 83.46$$

This chi-square statistic is similar to the one before except that there are 5 degrees of freedom. $\sum_{i=1}^{6} \frac{(n-e)^2}{e_i}$ where n and e have the same meaning as before.

This chi-square value of 83.46 is significant and, therefore, the hypothesis that the marginal totals in Table 2 are estimating the actual row totals is rejected. Note that the components for winter wheat and odd class are the major contributors to the significant chi-square.

There are no known statistical tests that compare one C.M. with another C.M. But there are two criteria that can be used to help evaluate a certain C.M. The first criterion simply assigns each misclassified point a loss of 1 and each correctly classified point a loss of 0. Under this criterion Table 1 has a loss value of 922 and Table 2 has a loss value of 654. This criterion is crude, but it seems reasonable for our purposes to give a misclassified corn pixel the same weight as the misclassified cotton pixel.

The next criterion is a bit more subtle. It uses the marginal totals in the C.M. For example, in Table 1, the column sum for cotton is 842. This means that 842 pixels were classified as cotton. Actually, there were 927 cotton pixels. In Table 2, there were 906 pixels classified into the cotton group. This is much closer to the correct number of 927. The marginal estimate (906) from Table 2 is within 2 percent of the actual. In Table 1, the marginal estimate is 842 or within 9 percent. Table 3 presents these estimates along with the percentages of the true value.

Table 3.--Marginal estimate and difference from actual values

Group	Actual	Unweighted prior probabilities			Weighted prior probabilities		
		Estimate	Difference	Percent	Estimate	Difference	Percent
Cotton	927	842	85	9.2	906	21	2.2
Corn	58	131	73	125.9	43	15	25.9
Soybeans	852	453	399	46.8	866	14	1.6
Grass	240	329	89	37.1	277	37	15.4
Winter wheat	85	346	261	307.1	27	58	68.2
Odd	55	116	61	110.9	98	43	78.2

In every case, unequal prior probabilities were superior to the equal prior probability model and in some cases, substantially so. For example, the number of corn pixels for Table 1 was 131 or 125.9 percent of the difference from the actual 58. The number of corn pixels for Table 2 is 43 or 25.9 percent of the difference from the actual 58 pixels. Soybeans, a very important item, also shows a significant improvement over the equal probability model. Actually, the soybean estimate for the equal prior probability model was 46.8 percent while the estimate for the unequal prior probability model was 1.6 percent.

Next, the point classification systems were compared to the per-field classification scheme. Table 4 presents the C.M. for the per-field classifier system.